Exploiting Image Semantics for Picture Libraries

Kobus Barnard
Computer Science Division
University of California, Berkeley
510-642-4979
kobus@cs.berkeley.edu

David Forsyth
Computer Science Division
University of California, Berkeley
510-642-9582
daf@cs.berkeley.edu

ABSTRACT

We consider the application of a system for learning the semantics of image collections to digital libraries. We discuss our approach to browsing and search, and investigate the integration both in more detail.

Categories and Subject Descriptors
H.3 [**Information Storage and Retrieval**].

General Terms
Algorithms, Performance, Human Factors

Keywords
Digital libraries, hierarchical image clustering

1. INTRODUCTION

Recently we have introduced a method for learning the semantics of collections of images from features and associated text˚[1]. In this paper we further explore the application of these ideas to digital image libraries. We consider the nature of the search and browsing processes, and argue that for many applications these should be used together. Specifically, search results should be organized by image semantics and visual structure.

This strategy is natural if the query is too general because browsing allows the user to refine the search based on a visual representation of the structure of the retrieved subset. The approach also makes sense if the query is too specific because a desired behavior in this case is to generalize the query in a sensible way, but there are typically several possible generalizations, and thus a mechanism for the user to choose an appropriate interpretation is needed.

2. THE MODEL

To model image data sets we use a generative statistical model initially developed for text [2], and extended for the joint occurrence of image features and associated text [1]. The images are considered to be composed of a number of segments and words. The occurrence of a segment (e.g. round orange disk) is analogous to the appearance of an associated word (sun).

The distribution of image items (words and segments) is generated by a hierarchical clustering model where each node in the tree emits words and image segments with node specific probability distribution. The emission probabilities are conditionally independent given a node, which roughly correspond to a hidden semantic or visual entity. To the extent that a document is in a specific cluster, it is generated by the nodes on the path to the route. The weighting of the vertical node distributions is document dependent.

3. QUERYING THE MODEL

We can query the image data set model with an arbitrary combination of words and representative image segments. A conjunction of query items is implicitly treated as a conjunction of hidden semantic and/or visual quantities. Because of this, and the probabilistic interpretation of the query process, we can get retrieve good images in response to queries which contain word combinations not present in any of the image annotations. For example, if no image in the database is associated with both the words river and tiger, the query river & tiger can still return images of tigers in water.

4. BROWSING QUERY RESULTS

In the river & tiger example, the word river is translated into a latent semantic quantity, which either approximates or generalizes river to yield success. However, the same process applies to tiger which may become jungle animals, and in conjunction with river yields a hippopotamus in a river. Thus the nature of the probabilistic query process can lead to ambiguous results. Ranking these results based on probability can lead to the presentation of many images corresponding to a less desirable alternative interpretation before any images corresponding to the desired interpretation appear. Thus to take full advantage of the power of our search approach, we propose to apply the clustering model to the search results in order to expose the available semantic and visual choices. Since the number of documents to be clustered is small, this clustering can be done fast enough for interactive searching, somewhat along the lines the scatter/gather approach [3, 4]. In addition to being much smaller, the tree for browsing may also have different fan out than that used for the complete dataset.

5. REFERENCES

[1] K. Barnard and D. Forsyth, Learning the Semantics of Words and Pictures, *Proc. International Conference on Computer Vision*, 2001, (in press).

[2] T. Hofmann, Learning and representing topic. A hierarchical mixture model for word occurrence in document databases, *Proc. Workshop on learning from text and the web*, CMU, 1998.

[3] M. A. Hearst and J. O. Pedersen, Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, *Proc. ACM SIGIR*, Zurich, 1996.

[4] F. Chen, U. Gargi, L. Niles, and H. Sch tze, Multimodal browsing of images in web documents, *Proc. SPIE Document Recognition and Retrieval*, 1999.